(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 August 2003 (14.08.2003)

PCT

(10) International Publication Number
WO 03/067466 A2

(51) International Patent Classification[7]: G06F 17/30

(21) International Application Number: PCT/IB03/00217

(22) International Filing Date: 21 January 2003 (21.01.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
02075498.2 5 February 2002 (05.02.2002) EP

(71) Applicant *(for all designated States except US)*: **KONINKLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) **Inventors; and**

(75) **Inventors/Applicants** *(for US only)*: **HAITSMA, Jaap, A.** [NL/NL]; Prof . Holstlaan 6, NL-5656 AA Eindhoven (NL). **KALKER, Antonius, A., C., M.** [NL/NL]; Prof . Holstlaan 6, NL-5656 AA Eindhoven (NL). **SCHIMMEL, Steven, M.** [NL/NL]; Prof . Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agent: **SCHMITZ, Herman, J., R.**; Internationaal Octrooibureau B.V., Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

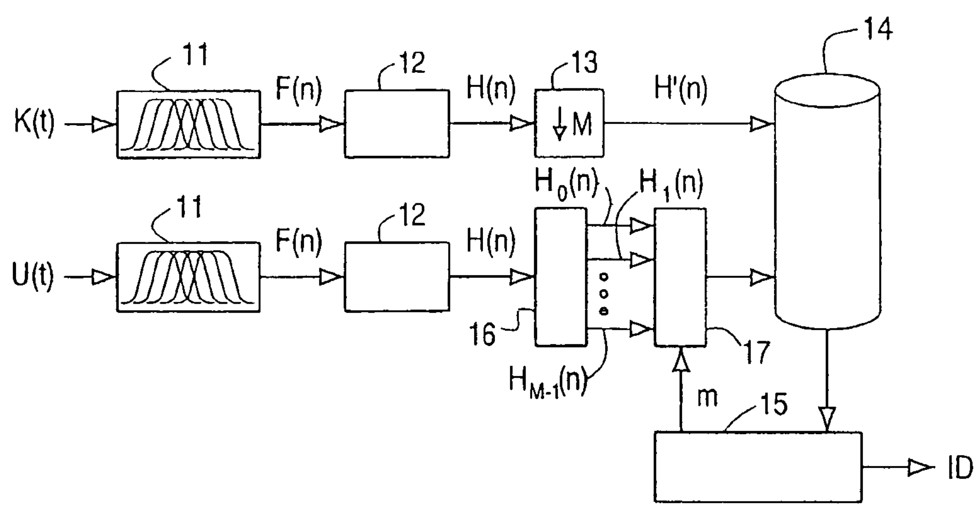
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: EFFICIENT STORAGE OF FINGERPRINTS



(57) **Abstract:** A robust fingerprinting system is disclosed. Such a system can recognize unknown multimedia content (U(t)) by extracting a fingerprint (a series of hash words) from said content, and searching a resembling fingerprint in a database in which fingerprints of a plurality of known contents (K(t)) are stored. In order to more efficiently store the fingerprints in the database and to speed up the search, the hash words (H(n)) of known signals (K(t)) are sub-sampled (13) by a factor M prior to storage in the database (14). The hash words (H(n)) of unknown signals (U(t)) are divided (16) into M interleaved sub-series (H0(n)..HM-1(n)). The interleaved sub-series are selectively (17) applied to the database (14) under the control of a computer (15). If only one of the sub-series sufficiently matches a stored fingerprint, the signal is identified.

WO 03/067466 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Efficient storage of fingerprints

FIELD OF THE INVENTION

The invention relates to a method and arrangement for storing fingerprints identifying audio-visual media signals in a database. The invention also relates to a method and arrangement for identifying an unknown audio-visual media signal.

BACKGROUND OF THE INVENTION

A fingerprint (in literature also referred to as signature or hash) is a digital summary of an information signal. In cryptography, hashes have been used for a long time to verify correct reception of large files. Recently, the concept of hashing has been introduced to identify multi-media content. Unknown content such as an audio or video clip is recognized by comparing a fingerprint extracted from said clip with a collection of fingerprints stored in a database. In contrast with a cryptographic hash, which is extremely fragile (flipping a single bit in the large file will result in a completely different hash), a fingerprint extracted from audio-visual content must be robust. To a large extent, it must be invariant to processing such as compression or decompression, A/D or D/A conversion.

A prior-art fingerprinting system is disclosed in Haitsma et al.: Robust Hashing for Content Identification, published at the Content-Based Multimedia Indexing (CBMI) conference in Brescia (Italy), 2001. As described in this article, the fingerprint is derived from a perceptually essential property of the content, viz. the distribution of energy in bands of the audio frequency spectrum. For video signals, the distribution of luminance levels in video images has been proposed to constitute the basis for a robust fingerprint.

A fingerprint is created by dividing the signal into a series of (possibly overlapping) frames, and extracting a hash word representing the perceptually essential property of the signal within each frame to obtain a respective series of hash words. In order to identify an unknown clip, the database receives the series of hash words concerned, and searches the most similar stored series of hash words. Similarity is measured by determining how many bits of the series match a series of hash words in the database. If the BER (Bit Error Rate, the percentage of the non-matching bits) is below a certain threshold, the clip is

identified as the song or movie from which the most similar series of hash words in the database originates.

A problem of the prior-art fingerprinting method is the size of the database. In the Haitsma et al. article, the audio signal is divided into frames of 0.4 seconds with an overlap of 31/32. This results in a new frame every 11.6 ms (=0.4/32). For every frame, a 32-bit hash word is extracted. Accordingly, a 5-minute song needs approximately 100 kbytes, viz. 5 (minutes) × 60 (seconds) × 4 (bytes per hash word) / 0.0116 (seconds per hash word). Needless to say that the database must have a huge capacity to allow recognition of a large repertoire of songs. Similar considerations apply to video fingerprinting systems.

OBJECT AND SUMMARY OF THE INVENTION

It is an object of the invention to provide a method and system for storing fingerprints in a database, which alleviates the above-mentioned problem. It is also an object of the invention to provide a method and system for identifying an unknown audio-visual signal in such a database.

To this end, the invention provides a method for storing fingerprints in a database as defined in independent claim 1. The method differs from the prior art in that only a sub-sampled sequence of hash words (i.e. one out of every M hash word) is stored in the database. The word "sequence" is used in this claim to refer to a full-length signal (song or movie). A storage reduction by a factor M is achieved.

A method of identifying an unknown audio-visual signal in such a database is defined in independent claim 4. As there is uncertainty as to which of M possible sub-sampled sequences of hash words is stored in the database, a full (i.e. not sub-sampled) series of hash words is extracted from the unknown clip in accordance with this method. The word "series" is used here to refer to a possibly short segment or clip of the unknown signal. Interleaved sub-series of hash words are now successively applied to the database for matching with the sub-sampled sequences stored therein. If at least one of the applied sub-series has a BER below a certain threshold, the signal is identified.

It is achieved with the invention that the storage requirements are reduced (by a factor M), while the robustness and the reliability of the prior-art identification method are maintained.

Further advantageous embodiments of the methods are defined in the dependent claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows a schematic diagram of an arrangement for storing and identifying fingerprints of audio-visual media signals in a database in accordance with the invention.

Fig. 2 is a diagram to illustrate a first operational mode of the arrangement which is shown in Fig. 1.

Fig. 3 is a diagram to illustrate a second operational mode of the arrangement which is shown in Fig. 1.

Fig. 4 is a flow chart of operational steps performed by a computer which is shown in Fig. 1.

DESCRIPTION OF EMBODIMENTS

The invention will be described for audio signals. Fig. 1 shows a schematic diagram of an arrangement in accordance with the invention. The arrangement is used for storing fingerprints of known audio signals in a database (first operational mode), as well as for identifying an unknown audio signal (second operational mode).

The first operational mode (storage) of the arrangement will be described first. In this mode, the arrangement receives a full-length music song K(t). The signal is divided, in a framing circuit **11**, into time intervals or frames F(n) having a length of approximately 0.4 seconds and weighted by a Hanning window with an overlap of 31/32. The overlap is used to introduce a large correlation between subsequent frames. For audio signals, this is a prerequisite because the framing applied to unknown signals to be recognized may be different.

The framing circuit **11** generates a new frame every 11.6 ms (=0.4/32). A hash extracting circuit **12** generates a 32-bit hash word H(n) for every frame. A practical embodiment of such a hash extracting circuit is described in the Haitsma et al. article referred to in the chapter Background of the Invention. Briefly summarized, the circuit divides the frequency spectrum of each audio signal frame into frequency bands and produces for each band a hash bit indicating whether the energy in said band is above or below a given threshold. Fig. 2 shows a sequence of hash words **21** thus obtained.

In accordance with the invention, the sequence of hash words is sub-sampled by a factor M by a sub-sampler **13**, which produces a sub-sequence H'(n). The sub-sequence of hash words, along with identification data such as title of the song, name of the artist, etc., constitutes a fingerprint of the known music song. Such a fingerprint is shown in Fig. 2,

where numeral **22** denotes the sub-sequence of hash words, and numeral **23** denotes title, artist, etc., identifying the song. The fingerprint is stored in a database **14** under the control of a computer **15**. In this example, where a sub-sampling factor M=4 has been used by way of example, a 5-minute song requires approximately 6,000 × 32 bits storage capacity. This is a saving of 75% as compared with the prior-art system where sub-sampling is not applied. In practice, the storage operation described above is performed for a huge number of known music songs. It will be appreciated that the order of the operations of hash word extraction (**12**) and sub-sampling (**13**) may be reversed.

The second operational mode (identification) of the arrangement will now be described. In this mode, the arrangement receives a part (say, 3 seconds) of an unknown song. i.e. an audio clip U(t). The clip is processed by a similar (or the same) framing circuit **11** and hash extracting circuit **12** as described above. The hash extraction circuit **13** extracts a full hash block (no sub-sampling) of the clip. For a 3-second clip, this operation yields a series of approximately 256 hash words H(n). Such a series of hash words representing the unknown audio clip is also referred to as hash block. In an alternative embodiment, the hash block has been extracted by a remote station and is merely received by the arrangement.

The hash block is applied to an interleaving circuit **16**, which divides it into M interleaved sub-series or sub-blocks $H_0(n)$, $H_1(n)$,..$H_{M-1}(n)$, where M is the same integer as used in the sub-sampler **13** described above. Fig. 3 illustrates the interleaving process for M=4. In this Figure, numeral **31** denotes successive hash words of the hash block, numeral **32** denotes sub-block $H_0(n)$, numeral **33** denotes sub-block $H_1(n)$, and numeral **34** denotes sub-block $H_{M-1}(n)$.

The sub-blocks are applied to respective inputs of a selection circuit **17**. Under the control of the computer **15**, the sub-blocks $H_0(n)$, $H_1(n)$,..$H_{M-1}(n)$ are successively applied to the database **14** for identification. If a series of hash words is found in the database, for which the bit error rate BER (i.e. the percentage of non-matching bits between said series and the applied sub-block) is below a certain threshold, the fingerprint comprising said series of hash words identifies the unknown audio clip.

Fig. 4 shows a flow chart of this identification process which is performed by the computer **15**. In a step **41**, an index m obtains an initial value 0. The index m is applied to the selection circuit **17** so that the first interleaved sub-block $H_0(n)$ of hash words is selected for identification. In a step **42**, the selected sub-block $H_m(n)$ is applied to the database. In a step **43**, it is checked whether a resembling series of hash words has been found in the database. The word "resembling" is understood to refer to the series of hash words having the

lowest BER provided that said BER, is less than a given threshold T. An actual example of a strategy of searching the most resembling series of hash words in the database is disclosed in the Haitsma et al. article mentioned before. Advantageous embodiments of search strategies are also proposed in Applicant's pending unpublished European patent applications 01200505.4 (PHNL010110) and 01202720.7 (PHNL010510).

If the BER is below the threshold, the audio clip has been identified. The title and performer of the song as stored in the database (**23** in Fig. 2) are then communicated to the user in a step **44**. If that is not the case, the index m is incremented (step **45**) so that another one of the interleaved sub-blocks is applied to the database. If all M interleaved sub-blocks have been searched without success (step **46**), the audio clip could not be identified. This outcome is communicated to the user in a step **47**.

It is achieved with the invention that the database capacity is reduced by a factor M. It should be noted that the same reduction can effectively be achieved by choosing a different frame overlap, viz. 7/8 in the present example. This is true as far as the first operational mode (storage) is concerned. However, if the same overlap of 7/8 without interleaving was chosen in the identification process, the robustness and reliability of the identification would be seriously affected. The invention resides in the concept of interleaving in the second operational mode (identification). It is achieved thereby that at least one of the interleaved sub-blocks is derived from a series of frames that substantially matches (in time) the series of frames from which the stored hash words have been derived. The identification process in accordance with the invention yields substantially the same robustness and reliability as the prior-art (non-interleaving) method with an overlap of 31/32. A mathematical background thereof will now be given.

When a sub-sampling with a factor M is applied and if the bits in a hash block are random i.i.d. (independent and identically distributed), the standard deviation of the BER increases by a factor $\sqrt{M}$. This implies that either the robustness and/or the reliability is/are affected. If the threshold on the BER is kept the same, then the robustness is unaffected but the reliability decreases. If on the other side the threshold is decreased by an appropriate amount, then the reliability stays the same but the robustness decreases.

However, the bits in a hash block of an audio-visual media signal have a large correlation along the time axis, which is introduced by the large overlap of the framing and inherent correlation in music. Therefore, the standard deviation s does not increase by the factor $\sqrt{M}$ when sub-sampling with the factor M is applied. Experiments have shown that, for small values of M, the standard deviation does not even increase significantly at all. In a

practical system without sub-sampling, the threshold on BER is set to 0.35. If sub-sampling by a factor M=4 is applied, then the threshold has only to be lowered to 0.342. Therefore, the decrease of robustness is insignificant, whilst the needed storage in the database has been decreased by a factor of 4. Furthermore, the time needed to search a hash database will decrease simply because there are 4 times fewer hash values in the database.

The search speed can even be further increased by refraining from applying a further sub-block to the database if one of the sub-blocks (generally the first) appears to have a BER which is larger than a further threshold (which is substantially larger than the threshold T). Because of the large correlation between sub-blocks (due to the frame overlap and inherent correlation in music), it is unlikely that another sub-block will have a significantly lower BER.

A robust fingerprinting system is disclosed. Such a system can recognize unknown multimedia content (U(t)) by extracting a fingerprint (a series of hash words) from said content, and searching a resembling fingerprint in a database in which fingerprints of a plurality of known contents (K(t)) are stored. In order to more efficiently store the fingerprints in the database and to speed up the search, the hash words (H(n)) of known signals (K(t)) are sub-sampled (13) by a factor M prior to storage in the database (14). The hash words (H(n)) of unknown signals (U(t)) are divided (16) into M interleaved sub-series ($H_0(n)..H_{M-1}(n)$). The interleaved sub-series are selectively (17) applied to the database (14) under the control of a computer (15). If only one of the sub-series sufficiently matches a stored fingerprint, the signal is identified.

CLAIMS:

1. A method of storing fingerprints identifying audio-visual media signals in a database, the method comprising, for each audio-visual signal, the steps of:
- dividing said audio-visual media signal into a sequence of frames;
- sub-sampling said sequence of frames by a factor M to obtain a sub-sampled sequence of frames;
- extracting, for each frame of said sub-sampled sequence of frames, a hash word representing a perceptually essential property of the signal within said frame, to obtain a respective sub-sampled sequence of hash words;
- storing said sub-sampled sequence of hash words as fingerprint in said database.

2. A method as claimed in claim 1, wherein said successive frames are overlapping.

3. An arrangement for storing fingerprints identifying audio-visual media signals (K(t)) in a database, the arrangement comprising:
- framing means (11) for dividing said audio-visual media signals into a sequence of frames;
- sub-sampling means (13) for sub-sampling said sequence of frames by a factor M to obtain a sub-sampled sequence of frames;
- means (12) for extracting, for each frame of said sub-sampled sequence of frames, a hash word (H(n)) representing a perceptually essential property of the signal within said frame, to obtain a respective sub-sampled sequence of hash words;
- a database (14) for storing said sub-sampled sequence of hash words as fingerprint in said database.

4. A method of identifying an unknown audio-visual media signal, the method comprising the steps of:
- dividing at least a part of the unknown audio-visual media signal into a series of frames;

- extracting, for each frame, a hash word representing a perceptually essential property of the signal within said frame, to obtain a respective series of hash words;
- dividing said series of hash words into M interleaved sub-series of hash words;
- successively applying said M sub-series to a database in which, for a plurality of multi-media signals, a sub-sampled sequence of hash words has been stored;
- identifying the unknown signal as the multi-media signal of which at least a part of the stored sub-sampled sequence of hash words substantially matches at least one of the M applied sub-series of hash words.

5. A method as claimed in claim 3, wherein said successive frames are overlapping.

6. An arrangement for identifying an unknown audio-visual media signal, the arrangement comprising:

- framing means (11) for dividing at least a part of the unknown audio-visual media signal (U(t)) into a series of frames;
- means (12) for extracting, for each frame, a hash word representing a perceptually essential property of the signal within said frame, to obtain a respective series of hash words;
- interleaving means (16) for dividing said series of hash words into M interleaved sub-series of hash words;
- selection means (17) for successively applying said M sub-series to a database in which for a plurality of multi-media signals, a sub-sampled sequence of hash words has been stored;
- computer means (15) for identifying the unknown signal as the multi-media signal of which at least a part of the stored sub-sampled sequence of hash words substantially matches at least one of the M applied sub-series of hash words.

7. A method of identifying an unknown audio-visual media signal, the method comprising the steps of:

- receiving, from a remote station, a series of hash words generated by dividing at least a part of the unknown audio-visual media signal into a series of frames, and extracting, for

each frame, a hash word representing a perceptually essential property of the signal within said frame;

– dividing said series of hash words into M interleaved sub-series of hash words;
– successively applying said M sub-series to a database in which, for a plurality of multi-media signals, a sub-sampled sequence of hash words has been stored;
– identifying the unknown signal as the multi-media signal of which at least a part of the stored sub-sampled sequence of hash words substantially matches at least one of the M applied sub-series of hash words.

8. A method as claimed in claim 5, wherein said successive frames are overlapping.

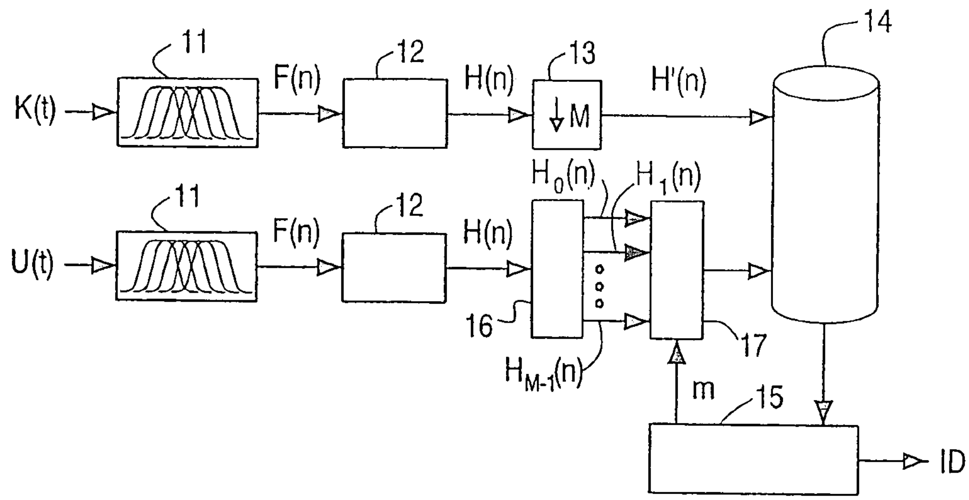

11
F(n)
12
H(n)
13
H'(n)
14
K(t)
M
11
12
$H_0(n)$
$H_1(n)$
U(t)
F(n)
H(n)
16
17
$H_{M-1}(n)$
m
15
ID

Fig.1

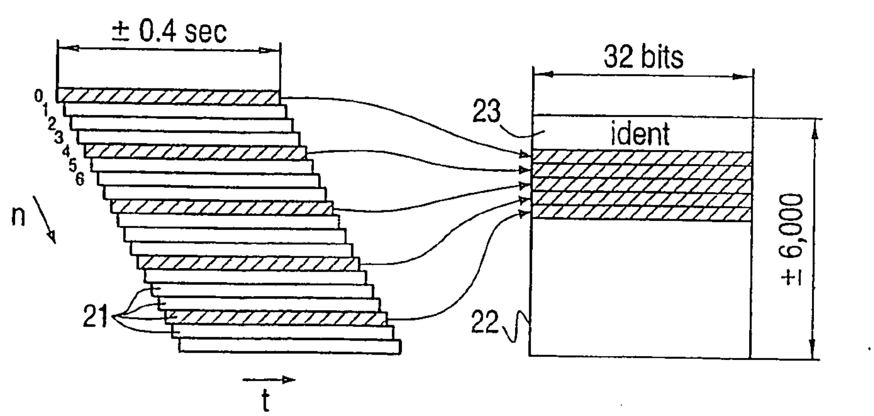

± 0.4 sec
32 bits
0
1
2
3
4
5
6
23
ident
n
± 6,000
21
22
t

Fig.2

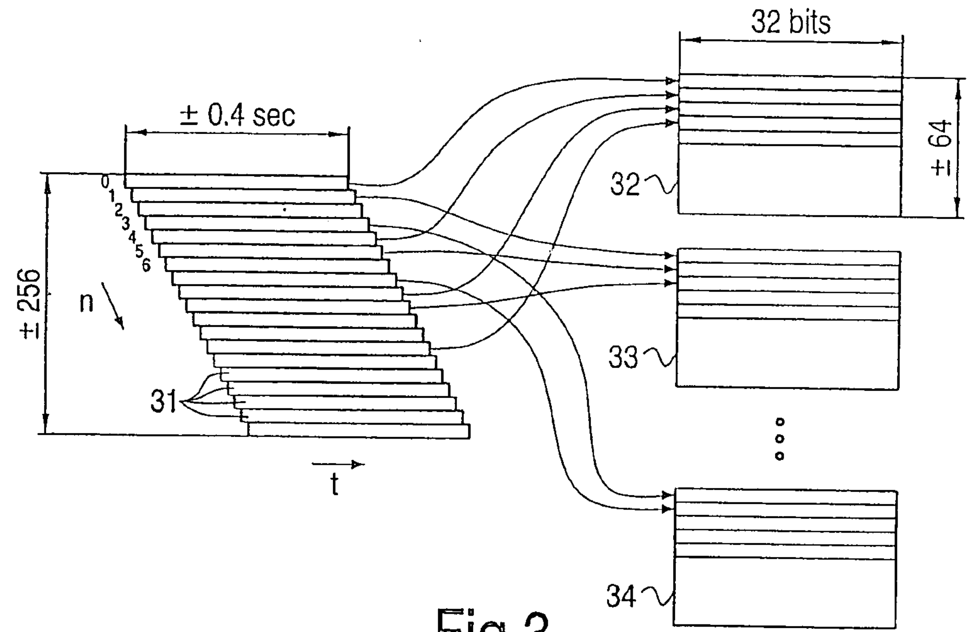

32 bits
± 0.4 sec
± 64
32
0
1
2
3
4
5
6
± 256
n
33
31
t
34

Fig.3

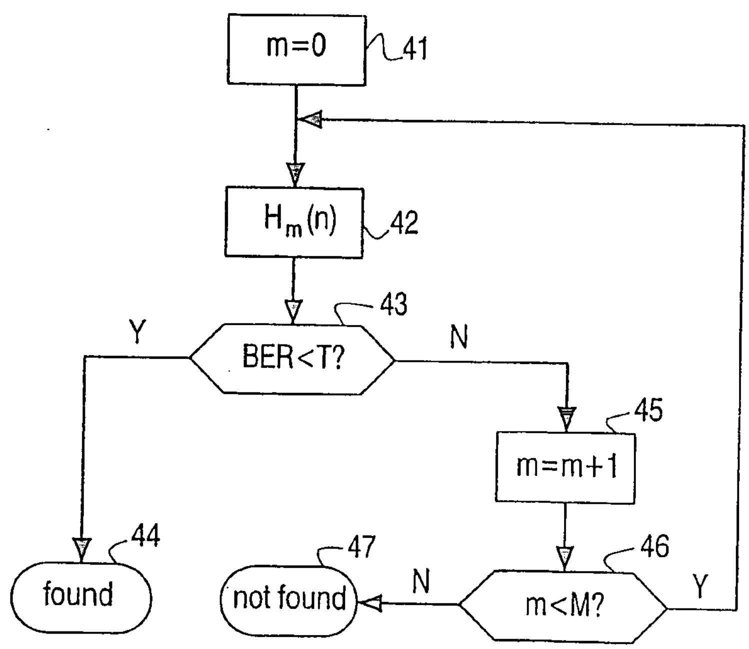

m=0
41
$H_m(n)$
42
43
Y
BER<T?
N
45
m=m+1
44
47
46
found
not found
N
m<M?
Y

Fig.4

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 August 2003 (14.08.2003)

PCT

(10) International Publication Number
WO 2003/067466 A3

(51) International Patent Classification[7]: G06F 17/30

(21) International Application Number: PCT/IB2003/000217

(22) International Filing Date: 21 January 2003 (21.01.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
02075498.2 5 February 2002 (05.02.2002) EP

(71) Applicant (*for all designated States except US*): **KONINKLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **HAITSMA, Jaap, A.** [NL/NL]; Prof . Holstlaan 6, NL-5656 AA Eindhoven (NL). **KALKER, Antonius, A., C., M.** [NL/NL]; Prof . Holstlaan 6, NL-5656 AA Eindhoven (NL). **SCHIMMEL, Steven, M.** [NL/NL]; Prof . Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agent: **SCHMITZ, Herman, J., R.**; Philips Intellectual Property & Standards, Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

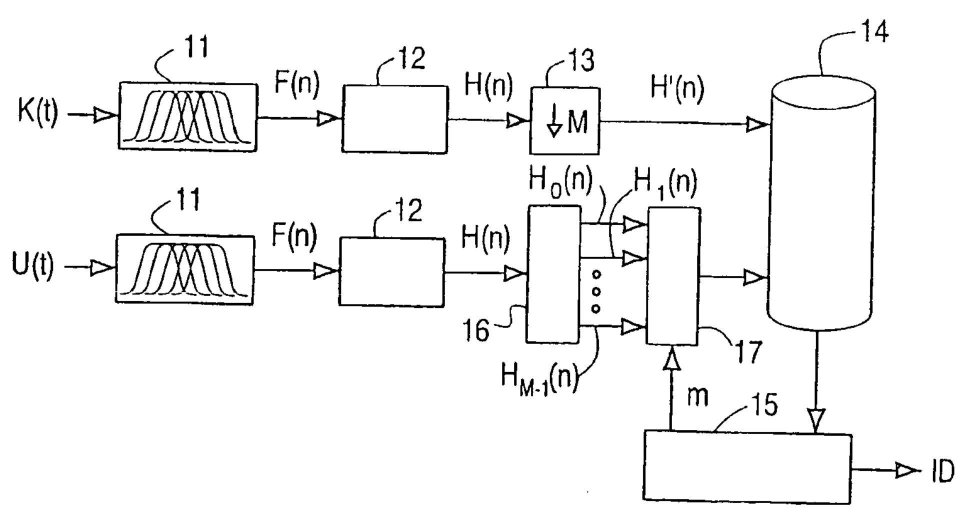
(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report: 3 June 2004

[Continued on next page]

(54) Title: EFFICIENT STORAGE OF FINGERPRINTS



(57) Abstract: A robust fingerprinting system is disclosed. Such a system can recognize unknown multimedia content (U(t)) by extracting a fingerprint (a series of hash words) from said content, and searching a resembling fingerprint in a database in which fingerprints of a plurality of known contents (K(t)) are stored. In order to more efficiently store the fingerprints in the database and to speed up the search, the hash words (H(n)) of known signals (K(t)) are sub-sampled (13) by a factor M prior to storage in the database (14). The hash words (H(n)) of unknown signals (U(t)) are divided (16) into M interleaved sub-series (H0(n)..HM-1(n)). The interleaved sub-series are selectively (17) applied to the database (14) under the control of a computer (15). If only one of the sub-series sufficiently matches a stored fingerprint, the signal is identified.

WO 2003/067466 A3

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/IB 03/00217

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06F H04N G11B H03M

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, INSPEC, PAJ, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HAITSMA ET AL: "Robust Audio Hashing for Content Identification" , PROCEEDINGS INTERNATIONAL WORKSHOP ON CONTENT-BASED MULTIMEDIA INDEXING, XX, XX, PAGE(S) 1-8 XP002198245 cited in the application page 1 -page 3 page 7 -page 8 ---	1-8
X	OOSTVEEN J ET AL: "VISUAL HASHING OF DIGITAL VIDEO: APPLICATIONS AND TECHNIQUES" , PROCEEDINGS OF THE SPIE, SPIE, BELLINGHAM, VA, US, VOL. 4472, PAGE(S) 121-131 XP001078972 ISSN: 0277-786X page 121 -page 124 --- -/--	1-8

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 February 2004	24/02/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Díaz de Lezana, C

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	SUBRAMANYA S R ET AL: "Transform-based indexing of audio data for multimedia databases" MULTIMEDIA COMPUTING AND SYSTEMS '97. PROCEEDINGS., IEEE INTERNATIONAL CONFERENCE ON OTTAWA, ONT., CANADA 3-6 JUNE 1997, LOS ALAMITOS, CA, USA,IEEE COMPUT. SOC, US, 3 June 1997 (1997-06-03), pages 211-218, XP010239191 ISBN: 0-8186-7819-4 page 213, paragraph 3 -page 215, paragraph 1 ---	1-8
A	SCHNEIDER M ET AL: "ROBUST CONTENT BASED DIGITAL SIGNATURE FOR IMAGE AUTHENTICATION" , PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP) LAUSANNE, SEPT. 16 - 19, 1996, NEW YORK, IEEE, US, VOL. VOL. 3, PAGE(S) 227-230 XP002090178 ISBN: 0-7803-3259-8 page 230; figure 8 ---	1-8
A	CHENG YANG: "MACS: music audio characteristic sequence indexing for similarity retrieval" IEEE WORKSHOP ON APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS, XX, XX, 21 October 2001 (2001-10-21), pages 123-126, XP010566890 page 124 -page 125 -----	1-8